סטטיסטיקה

משתנה אקראי

גודל שמאפיין איבר מסוים בקבוצת איברים מאותו סוג, ושעשוי להשתנות מאיבר לאיבר באופן אקראי.

משתנה אקראי מאופיין על ידי **שם**, **מספר** האיבר שאותו הוא מאפיין, **וגודל** (ערך).

שכיח

הערך הנפוץ ביותר, המופיע מספר הפעמים הגדול ביותר בסדרה.

ערכו של שכיח לאו דווקא מספרי: הוא יכול להיות מספר (כאשר איברי הסדרה הם מספרים), ויכול להיות גם כל מאפיין אחר (כגון שם עצם, צבע וכ"ד).

שכיחות

מספר המופעים של ערך (מאורע) מסוים בסדרה.

שכיחות יחסית

יחס של שכיחות המאורע לסה"כ מספר המאורעות בסדרה.

שכיחות יחסית באחוזים

ערך השכיחות המבוטא באחוזים.

טבלת שכיחויות יחסיות באחוזים

סכום ערכי השכיחויות היחסיות של כל המאורעות בסדרה שווה ל – 100%:

דוגמאות

X – מספר המופיע על-גבי אחת מכמה קוביות-משחק, N – מספר הקובייה.

$$X_2 = 5, X_1 = 6, X_3 = 2$$

משמע: על-גבי קובייה מס.1 הופיעה ספרה 6, על פני קובייה מס.2 הופיע מספר 5 ועל-גבי קובייה מס.3 הופיע 2.

1) בסדרת המספרים:

1, 1, 1, 2, 4, 5, 5, 5, 5, 7, 7, 10, 15

השכיח הוא <u>מספר 5</u>.

2) בסדרת שמות התלמידים:

אילנה, אייל, רועי, רותם, רותם, רותם, נאור, יוני, תמיד – השכיח הוא <u>*רותם*</u>.

מספר	1	2	4	5	7	10	15
שכיחות	3	1	1	4	2	1	1

המספר 5 מופיע 4 פעמים מתוך סה"כ 13 איברי הסדרה.

השכיחות היחסית שווה: $W(5)=\frac{4}{13}$

$$W(5)=\frac{4}{13}\approx 0.31=31\%$$

מספר	1	2	4	5	7	10	15
שכיחות ב-%	23	8	8	30	15	8	8

סה"כ: 23 + 8+8+30+15+8+8 = 100

סטטיסטיקה

התיאור הגרפי של נתונים

דרך אחת להציג ערכי המשתנה והתפלגות השכיחויות – היא טבלה. אולם ייצוג נתונים באמצעות הטבלה אינו מאפשר הסקת מסקנות מהירה.
דרך אחרת לתיאור הנתונים היא הייצוג הגרפי.

מצולע שכיחויות

אם ידועות השכיחויות M_i עבור ערכי המשתנה האקראי X_i, אזי אפשר לסמן את הנקודות $(X_i;\ M_i)$ בצירי הקואורדינטות ולחבר אותן באמצעות קטעים ישרים. לקו המקוטע שנוצר מכונים *מצולע שכיחויות*.

היסטוגרם

דיאגרמת מלבנים, כאשר רוחב כל מלבן מסמן טווח ערכי X_i וגובהו – גודל השכיחות המתאימה. שטח המלבן פרופורציוני למספר הפעמים שערכי המשתנה X מופיעים בסדרה.

דיאגרמת מקלות (מכונה גם דיאגרמת עמודות או **דיאגרמת מוטות)**

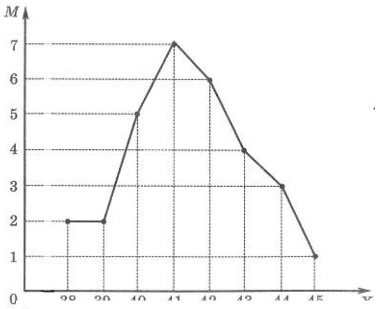
על ציר ה-X מופיעים ערכי המשתנים, ציר ה-Y הוא תדירויות הופעת משתנה בערך מסוים (שכיחות). על כל ערך אפשרי של המשתנה ניצב "מקל" בגובה פרופורציוני למספר הפעמים שערך זה מופיע בסדרה.

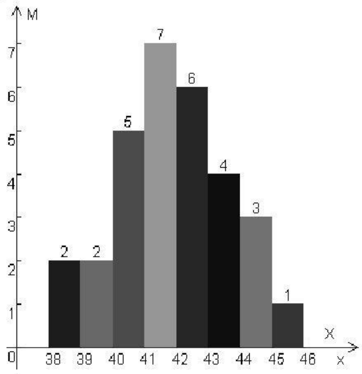
דוגמה

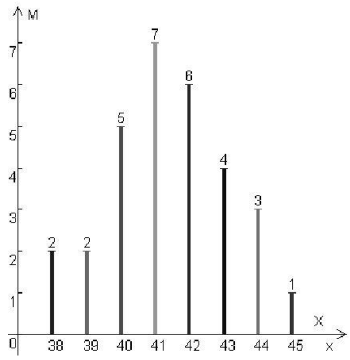
המשתנה האקראי X הוא מידת נעלי התלמידים של כיתה מסוימת. טבלת השכיחויות היא:

X	38	39	40	41	42	43	44	45
M	2	2	5	7	6	4	3	1

מצולע השכיחויות:



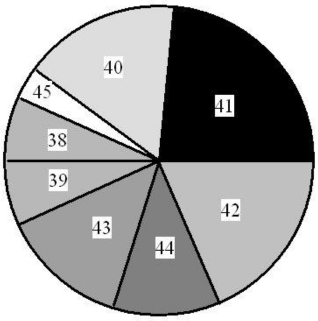




סטטיסטיקה

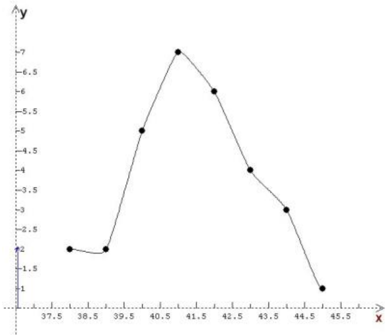
דיאגרמת עוגה

תרשים בצורת מעגל המציג התפלגות השכיחויות: שטח המעגל צבוע בצבעים (או גוונים) שונים, כפרוסות בעוגה, כאשר הזווית המוקדשת לכל צבע נמצאת ביחס ישר לשכיחות הופעת ערכי המשתנה שונים בסדרה.



גרף רציף

כאשר משתנה אקראי X מוגדר בתחום של מספרים ממשיים, מספר הנקודות במדגם שואף לאינסוף, ומצולע השכיחויות הופך לעקום חלק, המכונה *גרף התפלגות השכיחויות* (או *פיזור השכיחויות*).



מדדים סטטיסטיים

מדדי מרכז

מטרתם של מדדי מרכז היא אפיון מרכז ההתפלגות. ההבדל בין מדדי מרכז שונים הוא במידת רגישותם לערכים קיצוניים.

ממוצע

הערך ה"מרכזי" בתוך קבוצה של ערכים. **ממוצע חשבוני** $\overline{X}$ של קבוצת מספרים שווה לסכומם מחולק במספרם. אם כל ערכי הקבוצה שונים, אזי:

$$\overline{X} = \frac{X_1 + X_2 + \ldots + X_n}{N}$$

N – מספר כולל של הערכים.

עבור קבוצת הערכים 1,2,2,2,3,9 הממוצה החשבוני הוא:

$$\overline{X} = \frac{1+2+2+2+3+9}{6} = 3.17$$

מדדים סטטיסטיים

כאשר בין הערכים X_i מופיעים ערכים זהים בעלי שכיחות M_i, אזי נוסחת הממוצע החשבוני תהיה:

$$\overline{X} = \frac{X_1M_1+X_2M_2+ \dots + X_nM_n}{N}$$

כאשר: $N = M_1 + M_2 + \dots + M_n$

תכונות הממוצע החשבוני:

1) ערך הממוצע החשבוני אינו תלוי בסדר הופעת הנתונים;
2) אם נוסיף לכל איבר מספר כלשהו, ערך הממוצע החשבוני יגדל באותו המספר.
3) אם נחסיר מכל איבר מספר כלשהו, ערך הממוצע החשבוני יקטן באותו המספר.
4) אם נכפיל (נחלק) כל איבר באותו מספר, ערך הממוצע החשבוני יוכפל (יחולק) באותו מספר.

ממוצע הנדסי –

ממוצע הנדסי של מספר ערכים חיוביים הוא שורש מחזקת מספר הערכים ממכפלת הערכים:

$$\bar{x} = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n}$$

ממוצע הרמוני –

$$\bar{x} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \dots + \frac{1}{x_n}}$$

עבור סדרת הנתונים: 1,2,2,2,3,9

השכיחויות הן:

$M(1)=1, M(2)=3, M(3)=1, M(9) = 1$

והממוצע:

$$\overline{X} = \frac{1\cdot 1 + 2\cdot 3 + 3\cdot 1 + 9\cdot 1}{6} = 3.17$$

נחסיר מכל איבר שבסדרה 1,2,2,2,3,9 את המספר 2, ונקבל סדרה חדשה:

$-1, 0, 0, 0, 1, 7$

חישוב הממוצע:

$$\overline{X} = \frac{-1 + 1 + 7}{6} = \frac{7}{6} = 1.17$$

נוסיף 2, ונקבל תוצאה קודמת.

חישוב הממוצע החשבוני עבור הסדרה:

$0.1, 0.2, 0.2, 0.2, 0.3, 0.9$

נכפיל את כל האיברים ב- 10, ונקבל סדרת נתונים חדשה: 1,2,2,2,3,9.

עבורה: $\overline{X} = 3.17$. נחלק ב- 10, ונקבל את הממוצע המבוקש: $\bar{x} = 0.317$.

עבור $n = 2$: $\bar{x} = \sqrt{x_1 \cdot x_2}$

כאשר $x_1 = 2, x_2 = 10$:

הממוצע החשבוני: $\overline{X} = \frac{2 + 8}{2} = 5$

הממוצע ההנדסי: $\bar{x}=\sqrt{2\cdot 8} = \sqrt{16} = 4$

הממוצע ההרמוני:

$$\bar{x}=\frac{2}{\frac{1}{2}+\frac{1}{10}} = \frac{2\cdot 2\cdot 10}{2+10} = 3.33$$

מדדים סטטיסטיים

היחס בין הערכים הממוצעים:

עבור סדרת מספרים, הממוצע החשבוני שלהם תמיד גדול או שווה לממוצע ההנדסי, והלה גדול או שווה לממוצע ההרמוני שלהם:

$$\frac{x_1+x_2+\ldots+x_n}{n} \geqslant \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n} \geqslant \frac{n}{\frac{1}{x_1}+\frac{1}{x_2}+\ldots+\frac{1}{x_n}}$$

טווח –
המרחק בין הערך הגדול ביותר לבין הערך הקטן ביותר ($R = X_{max} - X_{min}$).

עבור סדרת הנתונים 1,2,2,2,3,9 הטווח הוא: $R = 9 - 1 = 8$.

חציון –
הערך שנמצא באמצע קבוצת הנתונים, כאשר היא מסודרת בסדר עולה, מהקטן אל הגדול.
כאשר מספר הנתונים זוגי, החציון הוא ממוצע שני הנתונים האמצעיים.

נתונות שתי סדרות נתונים:

(1) 3, 3, 3, 4, 4, 5, 5, 5, 8, 12

(2) 3, 3, 4, 4, 4, 4, 5, 6, 7.

מספר האיברים בסדרה (1) – זוגי ($N = 10$), לכן החציון שווה לממוצע חשבוני של שני הערכים המרכזיים, 4 ו-5:

$$\frac{4+5}{2} = 4.5$$

עבור הסדרה (2), החציון שווה לערך המרכזי (האיבר החמישי): 4.

סטייה –
הפרש בין ערך האיבר לבין הערך הממוצע של הסדרה: $\Delta_i = x_i - \bar{x}$.
הסטייה יכולה להיות הן חיובית והן שלילית.
סכום כל הסטיות בסדרה שווה לאפס.

עבור הסדרה (1):

$$\bar{x} = \frac{3 \cdot 3+4 \cdot 2+5 \cdot 3+8+12}{10} = 5.2$$

ערכי הסטיות:

-2.2, -2.2, -2.2, -1.2, -1,2, -.0.2,

-0.2, -0.2, 2.8, 6.8.

סכום כל הסטיות שווה ל- 0.

סטייה מוחלטת ממוצעת –
ממוצע חשבוני של הערכים המוחלטים של כל הסטיות.

סטייה מוחלטת ממוצעת:

$$\overline{\Delta} = \frac{3 \cdot |-2.2| + 2 \cdot |-1.2| + 3 \cdot |-0.2| + 2.8 + 6.8}{10} = 1.92$$

מדדים סטטיסטיים

שׁונות

מידת ריחוקם מהערך הממוצע. מחושב כממוצע ריבועי הסטיות מהממוצע:

$$\sigma^2=\frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+\ldots+(x_n-\bar{x})^2}{N}$$

סטיית תקן

השורש הריבועי של השונות.

יתרונה על השונות הוא בכך שהיא נמדדת ביחידות הנתון המקורי.

במונח "סטייה" מתכוונים למרחק בין ערך בקבוצה לבין הממוצע.

סטיית התקן היא אחד ממדדי הפיזור, והיא תמיד אי-שלילית.

$$\sigma=\sqrt{\sigma^2}$$

$$\sigma=\sqrt{\frac{(x_1-\bar{x})^2+(x_2-\bar{x})^2+\ldots+(x_n-\bar{x})^2}{N}}$$

תכונות של סטיית תקן

1) אם נוסיף לכל איבר מספר כלשהו, סטיית התקן לא תשתנה.

3) אם נחסיר מכל איבר מספר כלשהו, סטיית התקן לא תשתנה.

4) אם נכפיל (נחלק) כל איבר באותו מספר חיובי, שאינו שווה לאפס, ערכה של סטיית התקן תוכפל (תחולק) באותו מספר.

5) סטיית תקן מבוטאת באותן היחידות שבהן מוצגים הנתונים עצמם (להבדיל משׁונות, אשר מבוטאת בריבוע של היחידות שבהן מוצגים הנתונים).

עבור סדרת הנתונים (1):

$$\sigma^2=\frac{3\cdot2.2^2+2\cdot1.2^2+3\cdot0.2^2+2.8^2+6.8^2}{10}=$$

$$=7.16$$

נחשב את סטיית התקן עבור הסדרה (1):

$$\sigma=\sqrt{\sigma^2}=\sqrt{7.16}=2.68$$

אם נוסיף לכל איבר x_i מספר כלשהו, יגדל הממוצע $\bar{x}$ באותו המספר;

לכן הסטיות $(x_i-\bar{x})$ (ההפרשים בין הערכים לבין הממוצע) גם הן לא ישתנו, ויחד איתן לא תשתנה גם סטיית התקן.

אם נכפיל כל איבר באותו מספר, יוכפלו בו גם כל הסטיות; השונות יוכפל בריבוע המספר, וסטיית התקן – במספר עצמו.

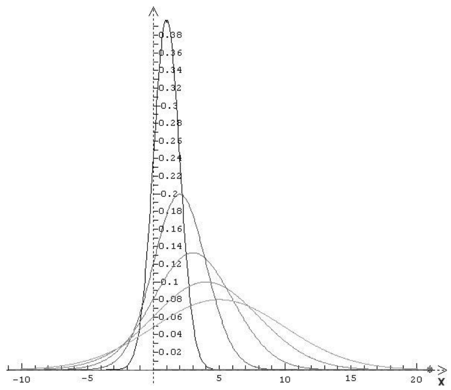
מדדים סטטיסטיים

התפלגות נורמלית

צורת עקומת השכיחויות של משתנים אקראיים רבים (כמו: תוצאות ניסוי, מדידות, מאפייני אוכלוסיה (גובה, משקל, מנת משכל ועוד)) היא **פעמון**, משמע: רוב הנתונים מתרכזים סביב ערך ממוצע באופן סימטרי;

הפעמון מתאפיין בשני הפרמטרים: מקום הערך הממוצע ו"רוחב" העקומה.

העקומה הנורמלית מתוארת על-ידי פונקצית גאוס:



$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

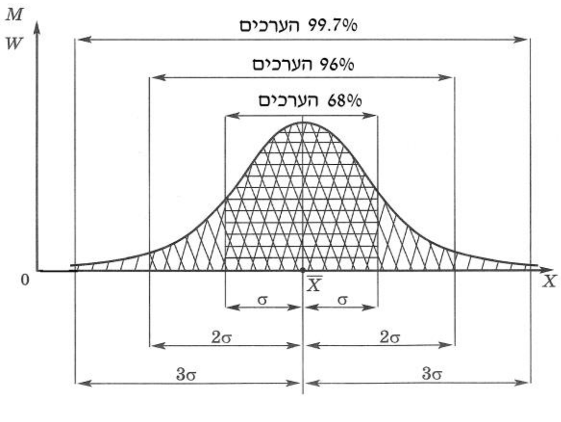
כאשר: σ - סטיית תקן,
μ - תוחלת.

התוחלת מייצגת את התוצאה הממוצעת אם חוזרים על ניסוי זהה פעמים רבות.

השטח הכלוא בין ציר ה- x לבין העקומה והאנכים $y = x_1$ ו- $y = x_2$ שווה לשכיחות (הסתברות ההתרחשות) של כל הערכים שבין x_1 ל- x_2. שטח מתחת לעקומת הפעמון $(-\infty < x < \infty)$ שווה ל- 1, ושטח מתחת לעקומה בין ציר – y ואינסוף שווה ל- 0.5.

חוק "שלוש σ" -

הערכים השכיחים ביותר מרוכזים במרכז הפעמון. קרוב ל- 68% מכל הערכים מרוכזים בתחום ± σ, ‏96% מכל הערכים מרוכזים בתחום ± 2σ, ‏99.7% מכל הערכים (למעשה כל הערכים!) מרוכזים בתחום של ± 3σ.



התפלגות נורמלית

<u>השטח הכלוא מתחת לעקומה נורמלית בין ∞- ל- x</u>

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

הערות:

בטבלה זו, x נמדד ביחידות של σ.

כדי למצא שטח הכלוא לדוגמה, בין 0 ל- σ, צריך מהערך עבור $x = 1$ להחסיר ערך עבור $x = 0$:

$$M(0,1) = M(1) - M(0) =$$

$$0.8413 - 0.5 = 0.3413 \approx 34\%$$

כדי למצא שטח הכלוא בין 2σ ל- 3σ, צריך מהערך עבור $x = 3$ להחסיר ערך עבור $x = 2$:

$$M(2,3) = M(3) - M(2) =$$

$$0.9987 - 0.9772 = 0.0215 \approx 2\%$$